EDUCATIONAL DATA FROM LARGE SCALE TESTING PROGRAMS: RESULTS AVAILABLE FROM COLLEGE LEVEL TESTING PROGRAMS

W. B. Schrader, Educational Testing Service

Since 1950, a number of salient changes have occurred in American higher education. A steadily increasing proportion of high school graduates has sought admission to colleges and universities. This trend, coupled with the marked increase in the birth-rate following World War II, has brought unprecedented demands for college education in recent years. As pressures for college admission increased, problems of college choice and needs for guidance became severe. Questions of "talent loss" and of needs for highly educated persons were widely discussed. Along a somewhat different line. there were many significant efforts to modify the high school curriculum led by outstanding scientists and scholars. As a result, the academic preparation of candidates for college admission has also been undergoing change. Finally, testing programs dealing with the transition from high school to college and from college to graduate and professional school have grown rapidly during this period.

This report will be concerned with data developed by college level testing programs and relevant to current questions in higher education. College level testing will be considered to include tests administered in the junior and senior years of high school if they are oriented toward planning for college as well as tests administered during the college years. It will be assumed that a testing program includes the development or selection of tests, their administration under standard conditions, scoring and reporting, and the development of materials (often statistical) to aid in score interpretation. In particular, the preparation of norms will be considered an integral part of a testing program, even if this involves special test administrations not part of the basic program operations.

Essentially, this report will be concerned with comparisons of various educationally significant groups with respect to performance on widely-used tests of academic ability and achievement. Brief treatment will be given to each of the following questions:

- -- How different in ability are college freshmen from high school seniors?
- -- How do major fields of study differ with respect to the ability of students that they attract?
- -- What indications of trends can be observed on the basis of year-to-year comparisons of student groups?

In addition, I plan to describe briefly a study now under way to learn more about the current state of subject-matter preparation of candidates for admission to college.

The relation of test performance to college attendance was studied on a state-wide basis by Learned and Wood (1938) in their classic Pennsylvania study of the 1930's. Other early state-wide studies by Toops (1940) in Ohio, Phearman (1949) in Iowa and Berdie (1954) in Minnesota supported the generalization that a large proportion of the highly able students were not attending college. Wolfle (1954) in his influential book <u>America's Resources of Specialized Talent</u> developed a national picture by judicious compilation of data from a variety of sources. Shortly afterward, a large national survey involving a brief test and a questionnaire and bearing on this question was initiated. This was the 1955 National Study of High School Students and Their Plans conducted by ETS as part of a broader study of scientific talent done by the College Board with the support of the National Science Foundation. Major reports emerging from this study include: Cole (1956), Stice, Mollenkopf, and Torgerson (1956), and Educational Testing Service (1957). More recently, Project Talent has developed a great quantity of information on this and many other topics. Among the many publications, Flanagan and others (1964) and Flanagan and Cooley (1966) are particularly relevant. Darley (1962) and a National Science Foundation report prepared by Bridgman have also made important contributions to the question of how ability is related to college-going. Like Wolfle. they used existing data to arrive at national estimates.

The utilization of data from testing programs to estimate the ability level of college freshmen is seriously handicapped by the fact that different colleges use different tests, and that some colleges presumably do not test their freshmen. Moreover, since only students who are considering enrollment in college are likely to take the College Board or the American College Testing Program tests, comparison of college freshmen with a cross-section of high school seniors cannot be obtained from operational data. One solution to this problem, clearly, is to administer a widelyused test to a representative sample of high school seniors and then follow them up to find out which ones enter college. This is essentially the NSF and Project Talent design except that it calls for using an operational test rather than spe-cially developed tests. Such a study has been done with the support of the College Entrance Examination Board to provide data to aid in score interpretation. This study utilized the Preliminary Scholastic Aptitude Test (PSAT) which is somewhat shorter than the Scholastic Aptitude Test (SAT) but is parallel in content. This parallelism is important because the SAT has been extensively validated as a predictor of college grades.

The initial purpose of the study was to develop high school norms for the PSAT and, indirectly, for the SAT. In drawing the sample of schools for these norms, a comprehensive list, maintained by ETS, of public and private secondary schools was used. Each school was classified on the basis of the nine census regions and as public or private. The number of schools to be drawn from each of the eighteen subsets was made proportional to the number of schools in the subset. Two hundred schools were selected for the base sample. Usable senior test data were obtained for 147 schools and 9,745 students. Although the proportion of schools cooperating was smaller than would be needed for definitive results, it was judged that the data were adequate to aid in score interpretation. Testing was conducted in October, 1960. The study is described in Chandler and Schrader (1966).

In the follow-up phase of the study, conducted by Seibel (1965), only a subsample of the 9,745 students was used. Using a composite score giving a weight of 2 to PSAT-Verbal and a weight of 1 to PSAT-Mathematical, all students in the top 5%, half of the students in the next 25% and one-tenth of the students in the bottom 70% were selected for follow-up. This reduced the number to 2,423. In the analysis, results for each student were weighted in accordance with the sampling ratio for his group. Some of Seibel's results are shown in Table 1. I must emphasize that these results describe not freshmen in general but those freshmen who entered college within a year after graduation from high school. There is good reason to believe that the means are somewhat higher than they would be for an entire freshman class. Subject to this limitation, however, they show the relative test performance of four educationally significant groups on a widely-used test. Rank-in-class as shown in Table 1 was converted to a percentile rank, and the resulting percentage expressed as a normal deviate on a scale having a mean of 13 and a standard deviation of 4. I might add that Seibel's data on rank-in-class were obtained from the high school and that the enrollment of each student in college was confirmed by the college. Seibel's figures for percent going to college within a year after high school graduation are: boys, 41%, girls, 32%, and total 36%. Project Talent results indicated that among the high school seniors in 1960, 46% of the boys, 33% of the girls, and 39% of the total reported that they had attended a two-year or four-year college since high school. However, 49% of the boys, 35% of the girls and 42% of the high school graduates stated that they had attended a recognized college, as reported by Flanagan and others (1964). The Project Talent results seem reasonably consistent with Seibel's findings.

From Seibel's report, it may also be estimated that, when PSAT 2V + M is used as the measure of ability;

- Of boys in the top 5% in ability, 96% went directly to college
- Of boys in the top 40% in ability, 74% went directly to college

- Of girls in the top 4% in ability, 91% went directly to college
- Of girls in the top 36% in ability, 66% went directly to college

It should be noted that the follow-up involved only 507 cases in the top 5%, and 1,240 cases in the next 25%, so that sampling error as well as sample bias must be considered in evaluating the percentages given. Nevertheless, they tend to support the view advanced by Berdie and Hood (1965) in their study of college plans of Minnesota high school seniors in 1960 that progress is being made in increasing the proportion of high ability youth going to college.

Scores earned in testing programs may yield useful information on ability differences among major fields of study and on career choices. An early example of relevant data is provided by Chauncey's (1952) article which included, along with extensive data on many aspects of the 1951 Selective Service College Qualification Test, data on the performance of students majoring in, or planning to major in, various subjects. These results indicated that, insofar as the sample was representative, a student enrolled in en-gineering, physical science or mathematics could expect to rank lower in ability relative to his fellow majors than a student of equal ability enrolled in, say, business or education. It should be clear, of course, that in any one college the relative ability levels might differ from the national pattern.

It should be noted that comparison of education majors with majors in other fields does not directly answer the question of the relative ability level of prospective teachers. In a study of college sophomores conducted in 1963 to provide norms for the Comprehensive College tests, it was possible to examine this question. In this study, as reported by Haven (1964), it turned out that of 1,227 prospective teachers who listed a major field, only 472 gave education as their major field. If prospective <u>college</u> teachers are excluded, 464 out of 1,046 prospective teachers, or less than half, gave education as their major field. When prospective teachers were classified by level of school, secondary school teachers were quite close to, though below, the general average in each of five tests covering various aspects of college work. In Natural Sciences and in Mathematics, they were about onefifth of a standard deviation below the general average. These results are based on about 150 cases.

Table 2 presents data drawn from the Graduate Examinations National Program with respect to Verbal and Quantitative scores earned by candidates in various fields. It may be useful in interpreting Table 2 to note that the scales for each test were established in 1952 so that a sample of college seniors would have a mean of 500 and a standard deviation of 100. All comparisons across fields are complicated, of course, by varying practices with respect to requirement of the tests by various departments and by national selection programs. Certainly no definitive conclusions can be drawn from them. They do illustrate, however, the possibility of developing information about major fields as a by-product of operational testing programs.

Test program data should be useful for making year-to-year comparisons of the ability levels of total candidate groups and of particular candidate subgroups. This statement assumes that a continuing score scale is maintained by appropriate equating methods. Most large-scale testing programs do have continuing scales to facilitate the use of scores. The utility of year-to-year comparisons should be greater if the candidate group represents a well-defined and stable population.

In my judgment, the Medical College Admis-sion Test has a particularly good set of statistics for year-to-year comparisons. The score scale was set in 1951 to yield a mean of 500 and a standard deviation of 100. Since then the scales have been maintained by equating. During this period, virtually all candidates for medical school admission took the test. Only one of the four tests has changed sufficiently to require a major change in title: "Understanding Modern Society" was replaced by "General Information" beginning in May 1962. A series of articles in the Journal of Medical Education has reported annual studies of the characteristics of medical applicants and accepted students. Thus a valuable body of data has been built up over the years. Data on number of applicants and the mean scores of applicants and accepted students are especially relevant. The complete series up to 1965 entrants can be found in two articles: Hutchins and Gee (1962) and Johnson (1965). On the Science test a consistent upward trend is apparent for classes admitted since 1957. The mean for applicants has risen from 482 to 513 and for accepted students, the mean has risen from 516 to 556 over the same period. It is more difficult to characterize the pattern in the other tests, except to note that the General Information test mean has been rising fairly noticeably since it was introduced and that mean scores in Verbal and Quantitative have been somewhat higher during the past five years than they were during the preceding five.

The Law School Admission Test, since it was instituted in 1948, has undergone a rapid growth both in the number of law schools requiring it and the number of candidates taking it. On the basis of unpublished data covering the testing years 1958-59 to 1964-65, there has been a consistent upward trend in scores from 1959-60 to 1964-65 from a mean of 477 to a mean of 510. (The standard deviation is about 100.) During this same period the candidate group nearly doubled in size -- from 20,735 to 39,162.

With respect to the College Board Scholastic Aptitude Test, the very rapid growth of this test both in the number of colleges requiring it and in the number of candidates taking it would seem to make year-to-year comparisons of scores difficult to interpret. From a different viewpoint, however, what is remarkable is the relatively small change in mean scores despite the marked change in the group tested. The mean Verbal score was 483 for seniors tested in the period December 1956 through March 1957 and 471 for seniors tested in the period December 1965 through March 1966. The corresponding means for Mathematical scores were 504 and 494. (For both tests the standard deviation for the more recent scores is about 110.) The number of candidates is more than four times as large for 1965-66 as for 1956-57. With respect to the verbal test, Stewart (1966) recently completed a study of the long-range stability of the score scale. She found that check equatings using December 1963 data agreed satisfactorily with original equatings done in March 1953 and in February 1957.

It may also be worth mentioning that boys scored 67 scaled score points higher on the mathematical test than did girls for the 1956-57 program data but only 46 points higher in the current year's data.

A somewhat different way in which data collected as an integral part of program operations may have some general usefulness is illustrated by the extensive questionnaire study of high school curricula now being conducted by ETS for the College Board. In this study, the aim is to describe the academic preparation of candidates currently choosing the various College Board achievement tests rather than of high school seniors or prospective college students generally.

The sample for this study included in all about 38,500 candidates. No candidate was asked to complete more than one questionnaire. The sampling plan allowed for slight overlap between fields, but a candidate drawn for two fields was assigned to the field having fewer candidates.

Ten questionnaires were prepared with the help of the examining committees in the various fields. The questionnaires cover English, Mathematics, History and Social Studies, three sciences -- Biology, Chemistry, and Physics -- and four languages -- Latin, French, German, and Spanish. Part of each questionnaire is concerned with the amount of work the student has taken in each subject, and with course titles in the particular field covered by his questionnaire, and in closely-related fields. Each questionnaire includes many questions about learning activities and course content by topics. For example, the mathematics questionnaire asks the student whether or not he has studied the normal curve, statistical sampling, and statistical inference.

In conclusion, data obtained in connection with operational testing programs may provide information relevant to broad educational questions. Test data collected in operational programs have two distinct advantages: students are likely to be highly motivated and relatively large samples can be obtained economically. For some purposes, however, the interpretation of results based on program data is difficult or impossible, particularly because the sample is self-selected. For such purposes, only a specially designed study can yield the information needed.

REFERENCES

- Berdie, Ralph F., <u>After High</u> <u>School</u> -- <u>What</u>? Minneapolis, <u>Minnesota</u>: The University of Minnesota Press, 1954.
- Berdie, Ralph F. & Hood, Albert B., <u>Decisions</u> for <u>Tomorrow: Plans of High School Seniors</u> for <u>After Graduation</u>. Minneapolis, Minnesota: University of Minnesota Press, 1965.
- Chandler, Marjorie Olsen & Schrader, William B., <u>Preliminary Scholastic Aptitude Test 1960</u> <u>National Norms</u>. Statistical Report 66-3. Princeton, New Jersey: Educational Testing Service, 1966.
- Chauncey, Henry, "The Use of the Selective Service College Qualification Test in the Deferment of College Students." <u>Science</u>, 1952, <u>116</u>, 73-79.
- Cole, Charles C., <u>Encouraging Scientific Talent</u>. New York: College Entrance Examination Board, 1956.
- Darley, John G., <u>Promise</u> and <u>Performance</u>: <u>A</u> <u>Study of Ability and Achievement in Higher</u> <u>Education</u>. Berkeley, California: Center for the Study of Higher Education, University of California, 1962.
- Educational Testing Service, <u>Background Factors</u> <u>Relating to College Plans and College En-</u> <u>rollment Among Public High School Students</u>. Princeton, New Jersey: Educational Testing Service, 1957.
- Flanagan, John C. & Cooley, William W., <u>Project</u> <u>Talent: One-Year Follow-Up Studies.</u> Cooperative Research Project No. 2333. Pittsburgh, Pennsylvania: School of Education, University of Pittsburgh, 1966.
- Flanagan, John C. and others, <u>The American High-School Student</u>. Cooperative Research Project No. 635. Pittsburgh, Pennsylvania: Project Talent Office, University of Pittsburgh, 1964.
- Harvey, Philip R. & Marco, Gary L., <u>Aptitude and</u> <u>Advanced Test Scores of 1963-1964 National</u> <u>Program Candidates by Undergraduate Major</u> <u>Fields.</u> Graduate Records Examination Special Report No. 65-3. Princeton, New Jersey: Educational Testing Service, 1965.
- Haven, Elizabeth W., <u>The Sophomore Norming Sample</u> for the <u>General Examinations of the Compre-</u> <u>hensive College Tests</u>. Princeton, New Jersey: Educational Testing Service, 1964.
- Hutchins, Edwin B. & Gee, Helen Hofer, "The Study of Applicants, 1960-61." <u>The Journal of</u> <u>Medical Education</u>, 1962, <u>37</u>, 1203-1212.

- Johnson, Doris G., "The Study of Applicants, 1964-65." <u>The Journal of Medical Education</u>, 1965, <u>40</u>, 1017-1030.
- Learned, William S., & Wood, Ben D., <u>The Student</u> <u>and His Knowledge</u>. Carnegie Foundation for the Advancement of Teaching Bulletin, No. 29. New York: The Foundation, 1938.
- National Science Foundation, <u>The Duration of</u> <u>Formal Education for High-Ability Youth</u>. Report NSF 61-36.
- Phearman, L. J., "Comparisons of High School Graduates Who Go to College with Those Who Do Not." <u>Journal of Educational Psychology</u>, 1949, <u>40</u>, 405-414.
- Seibel, Dean W., Follow-Up Study of a National Sample of High School Seniors: Phase 2 -One Year After Graduation. College Entrance Examination Board Research and Development Report, RDR 65-6, No. 1. Princeton, New Jersey: Educational Testing Service, 1965.
- Stewart, E. Elizabeth, <u>The Stability of the SAT-Verbal Score Scale</u>. Research Bulletin 66-37. Princeton, New Jersey: Educational Testing Service, 1966.
- Stice, Glen, Mollenkopf, Wm. G., & Torgerson, Warren S., <u>Background Factors and College-Going Plans Among Public High School Seniors</u>. Princeton, New Jersey: Educational Testing Service, 1956.
- Toops, Herbert A., "The Prediction of College-Going." <u>School and Society</u>, 1940, <u>51</u>, 257-261.
- Waite, Annette C. & Harvey, Philip R., <u>An Analysis</u> of <u>Graduate Record Examinations Scores by the</u> <u>Undergraduate Field of Study</u>, <u>1962-63</u>. Graduate Record Examinations Special Report 64-2. Princeton, New Jersey: Educational Testing Service, 1964.
- Wolfle, Dael, <u>America's Resources of Specialized</u> <u>Talent: A Current Appraisal and a Look Ahead</u>. New York: Harper and Brothers, 1954.

TABLE 1

Test Performance and Rank in Class for High School Seniors, College Entrants, Students Who Completed One Year of College, and Students Who Completed the Year in Good Standing^a

	PSAT-Verbal			PSAT-Mathematical			PSAT: 2V+M			Rank In Class		
Group	Mean	S.D.	мр	Mean	S.D.	Np	Mean	S.D.	мр	Mean	S.D.	мр
BOYS												
High School Seniors Entered College Completed One Year Completed Year in Good Standing	36.3 43.2 44.1 46.1	11.3 11.4 11.4 11.4	4585 1878 1599 1263	43.2 50.3 50.9 52.4	11.5 11.1 11.2 11.0	4585 1878 1599 1263	115.8 136.8 139.1 144.6	32.3 31.7 31.8 30.5	4585 1878 1599 1263	12.4 14.5 14.8 15.3	3.8 3.5 3.5 3.5	4199 1825 1558 1246
GIRLS												
High School Seniors Entered College Completed One Year Completed Year in Good Standing	37.0 46.0 46.6 47.3	11.8 12.2 11.8 11.7	5162 1659 1506 1377	38.5 45.9 46.4 47.0	10.3 10.5 10.5 10.3	5162 1659 1506 1377	112.5 137.9 139.6 141.7	32.2 32.6 31.8 31.2	5162 1659 1506 1377	14.3 16.3 16.4 16.6	3.6 3.4 3.4 3.3	4799 1609 1466 1349
TOTAL												
High School Seniors Entered College Completed One Year Completed Year in Good Standing	36.7 44.5 45.3 46.7	11.6 11.8 11.7 11.3	9747 3537 3105 2640	40.7 48.3 48.7 49.6	11.1 11.1 11.1 11.1	9747 3537 3105 2640	114.0 137.3 139.3 143.1	32.3 32.1 31.8 30.9	9947 3537 3105 2640	13.4 15.3 15.6 16.0	3.8 3.6 3.5 3.5	8998 3434 3024 2595

^aSource: Seibel, D. W. Follow-up Study of a National Sample of High School Seniors: Phase 2 - One Year After Graduation. College Entrance Examination Board Research and Development Reports, RDR 65-6, No. 1, 1965.

^bAll N's are weighted total N's.

Schrader, W. B. ASA, 8/16/66

TABLE 2

Mean Score on Graduate Record Examinations Aptitude Test for National Program Candidates by Major Field of Study

(Fields having 500 or more candidates in 1963-64)

		Mean					
	1962	2-63	1965	3-64	Number of Candidates		
Major Field	Verbal	Quanti- itative	Verbal	Quanti- itative	1962-63	1963-64	
Agriculture	419	507	421	505	553	785	
Biology	524	530	518	525	2086	3096	
Business	463	518	461	515	1775	2898	
Chemistry	554	643	551	640	2596	3557	
Economics	553	588	546	581	1334	2241	
Education	453	439	448	427	6266	8807	
Engineering	520	680	516	675	5432	7754	
English	599	493	600	492	4422	6953	
Fine Arts	529	455	524	450	415	645	
French	583	488	580	486	544	970	
Geology	542	592	549	599	604	717	
History	567	498	569	495	3206	5367	
Home Economics	438	421	432	410	317	551	
Humanities	603	529	609	530	521	703	
Mathematics	562	670	559	668	3153	4778	
Music	497	454	492	457	752	1086	
Natural Science	524	583	523	576	568	625	
Philosophy	624	562	621	565	751	1017	
Physical Education	402	417	399	411	773	1223	
Physics	589	697	589	695	2827	3472	
Political Science	584	531	582	528	1599	2476	
Psychology	565	531	565	529	3822	5696	
Religion	540	491	537	487	430	573	
Social Science	528	497	522	476	830	1252	
Sociology	533	486	532	477	952	1606	
Spanish	533	457	533	446	383	595	
Speech	507	440	501	433	372	551	
Zoology	542	548	546	560	771	1110	

^aSources: Waite, Annette C. and Harvey, Philip R., An Analysis of Graduate Record Examinations Scores by the Undergraduate Field of Study, 1962-63. Graduate Record Examinations Special Report 64-2, 1964.

Harvey, Philip R. and Marco, Gary L., Aptitude and Advanced Test Scores of 1963-64 National Program Candidates by Undergraduate Major Fields. Graduate Record Examinations Special Report 65-3, 1966.

^bNumbers of cases given are for the Verbal Score. Numbers for the Quantitative Score may differ slightly.